Abstract

The problem of how to select out of a large chemically accessible universe molecules representative of the diversity of that universe is resolved by the discovery of a method to validate molecular structural descriptors. Using the validated descriptors, optimally diverse subsets can be selected. In addition, from the universe, molecules with characteristics similar to a selected molecule can be identified. The validated descriptors also enable the generation of a huge virtual library of potential product molecules which could be formed by combinatorial arrangement of structural variations and cores. In this virtual library it is possible to search billions of possible product compounds in relatively short time frames.